SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

ORM

1. REPORT NUMBER	2. GOVT
AIM 1110	

AD-A214 718

4. TITLE (and Subtitle)

On the Verification of Hypothesized Matches in Model-Based Recognition

OVERED

memorandum

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

W. Eric L. Grimson and Daniel P. Huttenlocher

8. CONTRACT OR GRANT NUMBER(s)

N00014-86-K-0685
N00014-85-K-0124
DACA76-85-C-0010

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Artificial Intelligence Laboratory
545 Technology Square
Cambridge, MA 02139

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS

Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

12. REPORT DATE

May 1989

13. NUMBER OF PAGES

23

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

Office of Naval Research
Information Systems
Arlington, VA 22217

15. SECURITY CLASS. (of this report)

UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

object recognition
search
model-based vision

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Abstract. In model-based recognition a number of *ad hoc* techniques are used to decide whether or not a match of data to a model is correct. Generally an empirically determined threshold is placed on the fraction of model features that must be matched. In this paper we present a more rigorous approach in which the conditions under which to accept a match are derived based on fundamental grounds. We obtain an expression that relates the probability of a match occurring at random to the fraction of model features that are accounted for by the match. This expression is a function of the number of model cont.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1110 May, 1989

On the Verification of Hypothesized Matches in Model-Based Recognition

W. Eric L. Grimson
Daniel P. Huttenlocher[1]

Abstract. In model-based recognition a number of *ad hoc* techniques are used to decide whether or not a match of data to a model is correct. Generally an empirically determined threshold is placed on the fraction of model features that must be matched. In this paper we present a more rigorous approach in which the conditions under which to accept a match are derived based on fundamental grounds. We obtain an expression that relates the probability of a match occurring at random to the fraction of model features that are accounted for by the match. This expression is a function of the number of model features, the number of image features, and a bound on the degree of sensor noise.

One implication of our analysis is that a proper threshold for matching must vary with the number of model and data features. Thus it is important to be able to set the threshold as a function of a particular matching problem, rather than setting a single threshold based on experimentation. We analyze some existing recognition systems and find that our method yields thresholds similar to the ones that were determined empirically for these systems, providing evidence of the validity of the technique.

Acknowledgments: This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by an Office of Naval Research University Research Initiative grant under contract N00014-86-K-0685, and in part by the Advanced Research Projects Agency of the Department of Defense under Army contract number DACA76-85-C-0010 and under Office of Naval Research contract N00014-85-K-0124. WELG is supported in part by the Matsushita Chair of Electrical Engineering.

[1]Department of Computer Science, Cornell University, Ithaca NY

1. Introduction

A central problem in machine vision is that of recognizing partially occluded objects from noisy data. Recognition systems generally search for a matching between elements of an object model and instances of those elements in the data, recovering a transformation that maps part of the model onto part of the image. There are a number of different approaches to this model-based recognition problem, including clustering in parameter space (e.g., Stockman [1987], Stockman et al. [1982], Thompson and Mundy [1987]), searching a tree of corresponding model and image features (e.g., Grimson [1989a, 1989b] Grimson and Lozano-Pérez [1984, 1987], Ettinger [1987, 1988], Murray [1987a, 1987b], Murray and Cook [1988], Ayache and Faugeras [1986], Faugeras and Hebert [1986], Ikeuchi [1987]), and directly searching for possible transformations from a model to an image (e.g., Fischler and Bolles [1981], Huttenlocher and Ullman [1987, 1988]) (see also Chin and Dyer [1986] and Besl and Jain [1985] for more comprehensive reviews). These approaches all share the common property that a decision is made about the presence or absence of an object on the basis of geometric evidence acquired from the sensory input. In this paper we investigate the nature of this decision process and develop a formal means for deciding when a match should be accepted as correct.

To determine what constitutes an acceptable match of a model to an image, most recognition systems use one of two *ad hoc* approaches. The first approach is to find all possible interpretations and order them by some measure of completeness, such as the percentage of the model accounted for. The best interpretations, as defined by this measure, are then taken as correct solutions. Suppose one is looking for interpretations in the data of a particular object from the library of possible objects. If an instance of a particular object model is present in the scene and the measure of completeness is well behaved, then this approach will correctly find the interpretations. If no instance of a particular object model is present in the scene, the interpretations of this object that best account for the data are in fact incorrect. In this case, one must either accept false interpretations or there must be some means of deciding whether or not the object is present. Furthermore, this approach is computationally expensive, as in order to find all possible interpretations the entire search space must be accounted for.

The second common approach is to again apply a measure of completeness to each hypothesized match, but to use this measure to prematurely terminate the search as soon as an interpretation is found whose measure exceeds some threshold. Termination can be based strictly on the completeness of the current interpretation, or can involve examining the data for additional confirming or refuting evidence. Finding additional evidence can increase the measure of completeness of an interpretation, but one is still left with the problem of deciding whether an interpretation is good enough to accept as a correct match.

Current methods for deciding whether a match is correct are based on empirically determined thresholds. A more rigorous approach would be to derive conditions

under which to accept a match that are based on fundamental grounds. In this paper we analyze the problem of determining what constitutes a good match of a model to an image. In particular we derive an expression that relates the probability of a false match to the fraction of model features that are accounted for by the match. This expression is a function of the number of model features, the number of image features, and a bound on the degree of tolerable sensor noise. The derivation results from an examination of the likelihood of false positives (i.e., interpretations that are incorrect but arise due to a random coincidence of events in the image).

We then use this relation to define a threshold on the fraction of features that must be matched in order to limit the probability of a random coincidence to some level. We analyze some existing recognition systems ([Grimson and Lozano-Pérez, 1984, 1987] [Ayache and Faugeras, 1986] and find that our technique yields thresholds similar to the ones that were determined empirically for these systems. This provides experimental evidence of the validity of the technique, and suggests that it can be used profitably to set thresholds for other recognition tasks and systems.

Specifically, we address the following question:

- Suppose that we are given a model with m features, a set of s data features from a sensor, and bounds ϵ_p and ϵ_a on the positional and orientational error in the data. Further, suppose that some recognition method has found a match that accounts for a fraction f ($f \in [0,1]$) of the m model features. What is the relation between f and the likelihood δ that such a match can occur at random?

We use this relation to set a threshold on the minimum fraction of model features that must be matched, f_0, such that the likelihood of such a match occurring at random is small (e.g., $\delta < .001$). Note that there is not necessarily a value of f_0 for any choice of δ (in particular as δ gets very small, or as m, s, ϵ_p or ϵ_a get very large there may be no fraction of model features that limits the probability of a random match to δ).

There are three basic steps to the technique. First, given a particular type of feature, the type of transformation from a model to an image, and a bound on the sensor error, we characterize the set of transformations that are consistent with a single pairing of model and image features. This set of transformations is described by a volume V in the transformation space (a d-dimensional space with one dimension corresponding to each of the d transformation parameters).

We then determine the probability, $Pr\{e \geq l\}$ that the number of events (in this case the number of such volumes) that intersect at a common point in the transformation space is at least l. This likelihood is an estimate of how often a match of l features will occur at random. The probability of l volumes intersecting is estimated by considering the limiting case of a statistical occupancy problem as the number of observations and cells goes to infinity [Feller, 1968]. This method is similar to that used for the analysis of the generalized Hough transform in [Grimson and Huttenlocher, 1989].

Finally, the probability that l volumes will intersect at random is used to set a threshold on the minimum fraction of model features, f_0, that must be matched in order to accept an interpretation. In particular we set the threshold f_0 such that $mf_0 \leq l$, and $Pr\{e \geq l\}$ is a tolerable false matching rate δ.

2. The Space of Transformations

For rigid objects, the pose of an object with respect to a sensor can be characterized by a transformation from the model to the sensor coordinate systems. In this paper we focus on the case where this transformation is a similarity (i.e., consisting of translation, rotation, and scaling). The set of possible solutions to a given recognition problem can be viewed as a *transformation space* having one dimension corresponding to each parameter of the transformation from model to sensor coordinates. A point in this transformation space defines a pose of an object, which in turn defines a possible solution to the recognition problem. For example, with a two-dimensional image and world, the transformation space is four-dimensional (translation in x and y, rotation in the plane, and scaling).

A matching of a model feature with an image feature (such as an edge or vertex) defines a range of possible transformations from the model to the image, that is, a volume in the transformation space. The size and shape of this volume depends on the type of feature and on the degree of accuracy in the measurement of the features. In this section we present an analytic expression for the size of this volume. This expression is related to that developed in [Grimson and Huttenlocher, 1989] for characterizing the range of feasible transformations when the transformation space is tesselated at some sampling rate. Here, we determine an expression for the volume of feasible transformations in a continuous transformation space.

In this section we limit the discussion to the case of two-dimensional matching problems where the transformation is an isometry (translation and rotation without scaling), and the features are linear edge fragments. The method also applies to point features, as discussed at the end of the section. A similar analysis holds for three-dimensional matching problems and for problems involving change of scale, as described in the appendix.

Consider the problem of recognizing a two-dimensional polygonal model from noisy, occluded data. If $\mathcal{M}$ is the model coordinate system, we let

$\mathbf{M}_J$ be the vector to the midpoint of the J^{th} model edge, measured in $\mathcal{M}$,

$\hat{\mathbf{T}}_J$ be the unit tangent of the edge, measured in $\mathcal{M}$,

L_J be the length of the edge.

We let $\mathbf{m}_j, \hat{\mathbf{t}}_j, \ell_j$ denote similar parameters for the j^{th} data edge, measured in the sensor based coordinate system, $\mathcal{I}$. (Note that we use upper case characters to distinguish model parameters and lower case characters to distinguish sensory data parameters.)

The transformation from model coordinates to sensor coordinates may be represented by

$$\mathbf{v}_s = R_\theta \mathbf{V}_M + \mathbf{V}_0$$

where $\mathbf{V}_M$ is a vector in model coordinates, R_θ is a rotation matrix corresponding to an angle of θ, $\mathbf{V}_0$ is a translation offset, and $\mathbf{v}_s$ is the corresponding vector in sensor coordinates.

We need to know what transformations will map a model edge to a data edge. First, if $\ell_j > L_J$, we assume that the two edges cannot match. Thus, suppose that $\ell_j \leq L_J$. Then the rotation needed to align the two tangents is given by the angle θ_m between $\hat{\mathbf{T}}_J$ and $\hat{\mathbf{t}}_j$, and this defines a rotation matrix R_{θ_m}. If we apply this rotation to the set of edge points

$$\left\{ \mathbf{M}_J + \gamma \hat{\mathbf{T}}_J \quad \middle| \quad \gamma \in \left[-\frac{L_J}{2}, \frac{L_J}{2} \right] \right\}$$

we get a set of transformed points

$$\left\{ R_{\theta_m} \left[\mathbf{M}_J + \gamma \hat{\mathbf{T}}_J \right] \quad \middle| \quad \gamma \in \left[-\frac{L_J}{2}, \frac{L_J}{2} \right] \right\}.$$

To align the edges, we need to translate these rotated points. Now, because $\ell_j \leq L_J$, there are many transformations that will cause the edges to overlap. Consider one endpoint of the data edge

$$\mathbf{p}_1 = \mathbf{m}_j - \frac{\ell_j}{2} \hat{\mathbf{t}}_j.$$

If this happens to coincide with a model edge endpoint,

$$\mathbf{P}_1 = \mathbf{M}_J - \frac{L_J}{2} \hat{\mathbf{T}}_J$$

then

$$\mathbf{m}_j - \frac{\ell_j}{2} \hat{\mathbf{t}}_j = R_{\theta_m} \left[\mathbf{M}_J - \frac{L_J}{2} \hat{\mathbf{T}}_J \right] + \mathbf{V}_0$$

so that the translation is

$$\mathbf{V}_0 = \mathbf{m}_j - R_{\theta_m} \mathbf{M}_J + \frac{L_J - \ell_j}{2} R_{\theta_m} \hat{\mathbf{T}}_J$$

because $R_{\theta_m} \hat{\mathbf{T}}_J = \hat{\mathbf{t}}_j$. Similarly, if the other endpoints align, we get

$$\mathbf{V}_0 = \mathbf{m}_j - R_{\theta_m} \mathbf{M}_J - \frac{L_J - \ell_j}{2} R_{\theta_m} \hat{\mathbf{T}}_J.$$

Because any intermediate position is also acceptable, the set of translations consistent with matching model edge J to data edge j is given by

$$\left\{ \mathbf{m}_j - R_{\theta_m} \mathbf{M}_J + \gamma R_{\theta_m} \hat{\mathbf{T}}_J \quad \middle| \quad \gamma \in \left[-\frac{L_J - \ell_j}{2}, \frac{L_J - \ell_j}{2} \right] \right\}. \tag{1}$$

Hence, matching model edge J to data edge j yields a set of points in transform space $\mathcal{P}$, with a single value for the rotation parameter and a set of values for the translation, that correspond to a line of length $L_J - \ell_j$, with orientation $R_{\theta_m} \hat{\mathbf{T}}_J$ in the x–y plane. This is shown in Figure 1.

This, however, ignores the issue of noise in the measurements. In practice, we may only know the position of the endpoints of the data edge to within some ball

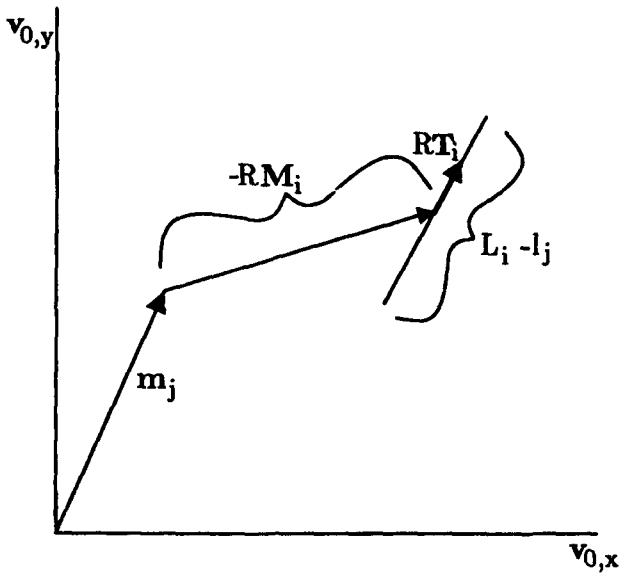

Figure 1. Range of feasible translations, for fixed θ and with no position error. The line in the direction of $R\mathbf{T}_i$ denotes the set of feasible translations for a given value of θ.

(which in two dimensions is just a circle) of radius ϵ_p, and the orientation to within an angular error of ϵ_a. For the case of two dimensional lines, these error ranges are related. Given endpoint variations of ϵ_p, it is straightforward to show that the maximum angular variation occurs when the correct line is tangent to both circles of radius ϵ_p about the two endpoints, and is given by

$$\epsilon_a = \tan^{-1}\left(\frac{2\epsilon_p}{\sqrt{\ell^2 - 4\epsilon_p^2}}\right)$$

provided $\ell > 2\epsilon_p$.

Inclusion of error effects on position measurements imply that the line of feasible translations, for a given rotation, (as given by equation (1)), must be expanded to include any points in the parameter space within ϵ_p of that line. Further, this expansion into a region must be repeated for each value of θ in $[\theta_m - \epsilon_a, \theta_m + \epsilon_a]$. Note that this carves out a skewed volume in transform space, because the region's center and orientation are functions of θ (see equation (1)). This observation has been carefully analyzed in [Clemens, 1986]. The volume is illustrated in Figure2.

Thus, given $\mathbf{M}_J, \hat{\mathbf{T}}_J, L_J, \mathbf{m}_j, \hat{\mathbf{t}}_j, \ell_j$, we will use the following conditions:

- If $\ell_j - 2\epsilon_p > L_J$, then there are no consistent transformations,
- Otherwise, the set of feasible transformations is denoted by the volume

$$\mathcal{V}(j,J) = \bigcup_{\theta \in [\theta_m - \epsilon_a, \theta_m + \epsilon_a]} \mathcal{S}(\theta, j, J)$$

where an individual set of translations is denoted by:

$$\mathcal{S}(\theta, j, J) = \left\{ (\theta, \mathbf{V_0}) \middle| \exists \gamma, |\gamma| \leq \frac{L_J - \ell_j}{2}, \|\mathbf{m}_j - R_\theta \mathbf{M}_J + \gamma \hat{\mathbf{T}}_J - \mathbf{V_0}\| \leq \epsilon_p \right\}.$$

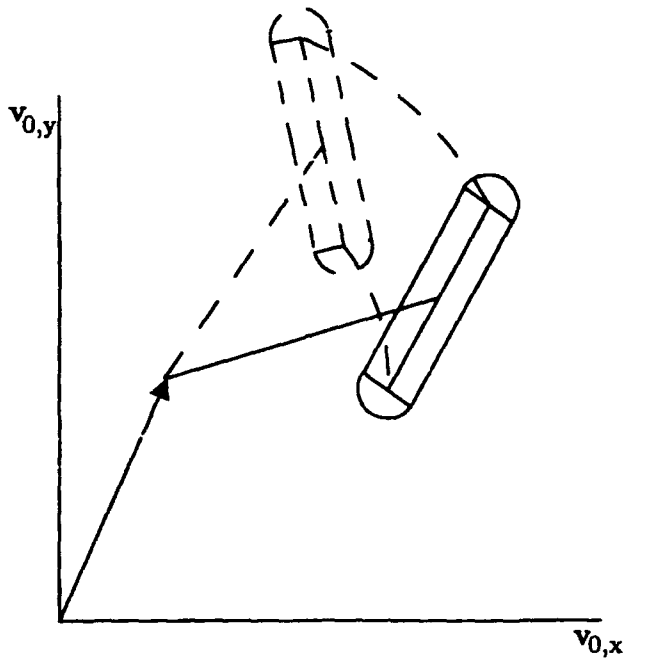

Figure 2. Range of feasible translations, with error. The region enclosed in solid lines indicates the slice $\mathcal{S}(\theta, j, J)$ for a particular value of θ. As θ varies, this slice rotates through a helical path, as indicated by the region enclosed by a dashed line.

We can use this expression to determine the size of the set of feasible transformations. Since each slice $\mathcal{S}(\theta, j, J)$ consists of two hemicircles and a rectangle, it is easy to show that the volume of the region defined above is given by

$$c_{jJ} = 2\epsilon_a \left[2\epsilon_p (L_J - \ell_j) + \pi \epsilon_p^2 \right].$$

The term in square braces corresponds to the area of a single slice, this is integrated over a range of angles, yielding the $2\epsilon_a$ term.

For simplicity, we let the data edge have a length

$$\ell_j = (1 - \alpha_{jJ}) L_J$$

where α_{jJ} denotes the amount of occlusion of the edge, so that the expression for the volume becomes

$$c_{jJ} = 2\epsilon_a \left[2\epsilon_p \alpha_{jJ} L_J + \pi \epsilon_p^2 \right]. \tag{2}$$

If we are dealing with point features, rather than extended edges, the above result can be specialized. Here $L_j \to 0$ so that equation (2) becomes

$$c_{jJ} = 2\epsilon_a \pi \epsilon_p^2. \tag{2a}$$

Now, what is ϵ_a in the case of a point feature? If the feature is a vertex, one can use the direction of the bisector of the two edges defining the vertex as the orientation of the vertex, and hence can bound the error in measuring that orientation as ϵ_a. Similarly, if the vertex is a curvature extremum or a point of curvature inflection, one can use the local tangent of the curve to define the orientation, and ϵ_a is again defined by a bound on measuring this orientation. If the vertices are truly isolated points, then $\epsilon_a = \pi$. In any event, our analysis provides estimates for c_{jJ} both for the case of edge features and for the case of vertices.

For the case of a rigid two-dimensional isometric transformation, we have characterized the volume of transformation space, c_{jJ} that is consistent with a single data-model pairing (j, J). This expression is given by equation (2) for edge features and equation (2a) for point features. The expression is a function of the noise in the data measurements, ϵ_p and ϵ_a, and in the case of edges is further a function of the amount of occlusion, α_{jJ}, and the length of the model edge, L_J. In the appendix we consider adding scaling to the transformation as well as the case of three-dimensional transformations. We now turn to the question of how these volumes interact.

3. The Probability of a Conspiracy

In the previous section, we characterized the volume of transformation space that is consistent with a data-model pairing. If two such volumes overlap, then their intersection defines the set of transformations that are consistent with both of the data-model pairings. Thus a correct match of a model to an image will lie in the intersection of several volumes. In this section we investigate the likelihood that l volumes in transformation space will intersect at random. Such an event corresponds to an arrangement of image features that happens to be consistent, within error bounds, with l of the model features, but which does not actually correspond to an instance of the object.

The likelihood that l transformation space volumes will intersect at random is a function their number and size. The number of volumes depends on the number of model and image features. The size of each volume depends on the amount of noise in the data, the type of feature, and for edge features the amount of occlusion of the edges. In order to be confident that a match accounting for l model features is correct, we would like to choose l such that the likelihood of a random matching of that size is very small.

In order to characterize the likelihood that several volumes will intersect at random we make use of a statistical occupancy model. In the discrete case, if r events are uniformly randomly distributed across n buckets, an occupancy model can be used to estimate the probability that a given bucket will contain k events. The events in our case are points in the volumes in transformation space, and the buckets are points in the transformation space itself. These events and buckets are continuous rather than discrete, and thus we are concerned with the limiting case as $n, r \rightarrow \infty$.

The volume of transformation space defined by each incorrect model and image feature pairing is independent of the correct match. Furthermore, we assume that the image features are independent of one another. Thus we can model the volumes in transformation space as independent random events. The distribution of these volumes depends on the image features, which are unknown, so we assume the uniform distribution as an approximation.

While the volumes in transformation space can reasonably be viewed as independent random events, we are modeling the probability of events occurring at points in these volumes. As the number of volumes, R, gets large (compared with the ratio of the total size of the transformation space to the size of each volume, V/c) the overall distribution of points in the space is also random. For the cases of interest here $Rc \gg V$, so the assumption of independent random pointwise events is a reasonable approximation.

An alternative explanation of the independence of the pointwise events in transformation space is the following. The probability that a particular point is consistent with a given data-model pairing is equivalent to the probability that the point lies within some neighborhood of the centroid of the given volume in transformation space. Since the image features are assumed to be independently randomly distributed, this probability is independent of the choice of image feature. Thus in the following analysis we assume that the events in transformation space are randomly distributed, and use the uniform distribution as an approximation.

Given a uniform random distribution of r events into n cells such that each of the n^r placements have equal probability, the probability that a given cell contains exactly k events is given by the binomial distribution

$$p_k = \binom{r}{k} \frac{1}{n^k} \left(1 - \frac{1}{n}\right)^{r-k}.$$

In the limit, as $n, r \to \infty$, where the ratio $\frac{r}{n} \to \lambda$, the binomial distribution is approximated by

$$p_k \approx \frac{\lambda^k}{k!} e^{-\lambda}.$$

This distribution is often termed the Maxwell-Boltzmann statistic (for a standard reference see [Feller, 1968]).

In addition to the Maxwell-Boltzmann distribution, another common distribution used in occupancy problems is the Bose-Einstein statistic, which has an experimental basis in particle physics. Under the Bose-Einstein model, for large r and n where $\frac{r}{n} \to \lambda$, the limiting case is the geometric distribution, where

$$p_k \approx \frac{\lambda^k}{(1+\lambda)^{k+1}}.$$

This distribution has a long tail as $k \to \infty$, and thus predicts large peaks with a higher probability than does the Maxwell-Boltzmann model. We are interested in establishing conservative bounds on the likelihood that a large number of volumes will intersect at random, thus we use the Bose-Einstein statistic because it provides a higher estimate of this likelihood.

The parameter λ of the occupancy model is the ratio of the occupied volumes of the transformation space to the total size of the transformation space. From equations (2) and (2a) in the previous section we know that each pair of model and image features defines a volume of size c_{jJ} in transformation space. There are ms such volumes for m model features and s image features, so the occupied volume of the transformation space is given by

$$\sum_{j=1}^{s}\sum_{J=1}^{m} c_{jJ}.$$

The total size of the transformation space is just the product of the ranges for the dimensions of the space. Each rotational dimension ranges over the interval $[0, 2\pi]$, and each translational dimension ranges over $[0, D]$, where D is the linear extent of the image. Thus in the case of a two dimensional isometry (translation and rotation) we get

$$\lambda = \frac{\sum_{j=1}^{s}\sum_{J=1}^{m} c_{jJ}}{2\pi D^2}.$$

We can simplify this to

$$\lambda = ms\overline{c}$$

where $\overline{c}$ is the average normalized volume size.

In the case of two-dimensional edge features, from equation (2) we obtain the average normalized volume size

$$\overline{c} = \frac{2\epsilon_a\left[2\epsilon_p\overline{\alpha}\overline{L} + \pi\epsilon_p^2\right]}{2\pi D^2}$$

where $\overline{L}$ is the average edge length, $\overline{\alpha}$ is the average amount of occlusion of the edges (the average value of α_{jJ}), and where we have incorporated the normalizing term $2\pi D^2$. Note that as expected $\overline{c}$ increases as the noise ϵ_a, ϵ_p increases, and also $\overline{c}$ increases as the average amount of occlusion of the edges $\overline{\alpha}$ increases.

For two-dimensional points features (with associated orientations), the average normalized volume size is given by

$$\overline{c} = \frac{2\epsilon_a\pi\epsilon_p^2}{2\pi D^2}.$$

Note that we can restrict $\epsilon \leq \pi$ and $\epsilon_p \leq \frac{D}{2}$. In the extreme case, this can lead to $\overline{c} > 1$, which does not make physical sense. We should really take the minimum of the above expressions and 1, but in practice $\overline{c}$ is usually much smaller than this and hence we ignore this special case.

A particular recognition task thus defines a value for λ, based on the type of transformation from the model to the image, the type of features, the number of model features m, the number of data features s, and a bound on the positional and angular error, ϵ_p and ϵ_a.

Given a value for λ, we are interested in the probability that l or more of the volumes intersect at random, which is given by

$$Pr\{e \geq l\} = 1 - \sum_{k=0}^{l-1} p_k.$$

This corresponds to an arrangement of data features occurring at random that such that l of the model features can be matched (within the error bounds) to those data features. From $Pr\{e \geq l\}$ we can determine the fraction of model features, f_0, such that the probability of mf_0 features being matched at random is less than some predefined level, δ. This value is just the smallest f such that

$$Pr\{e \geq mf\} \leq \delta$$

i.e.

$$f_0 = \min\{f \mid Pr\{e \geq mf\} \leq \delta\}.$$

We then choose δ such that the probability of a false match is small, for example $\delta = .001$.

The analysis in this section simply counts each pairing of a data feature with a model feature equally. It is also possible to weight the events by the amount of model accounted for. Below we consider the case of weighting each feature match by the length of the matched edge.

4. Deriving Formal Thresholds

We have used an occupancy model to determine an expression for the probability that l or more volumes in transformation space will intersect at random. This expression is a function of the number of features, the type of features, and bounds on the sensor error. The expression was then used to set a threshold, f_0, on the fraction of model features that must be matched in order to limit the probability of a random matching to some level. In this section we derive a closed-form expression for f_0.

Under Bose-Einstein statistics, we have

$$p_k \approx \frac{\lambda^k}{(1+\lambda)^{k+1}}$$

or equivalently

$$p_k \approx \frac{1}{1+\lambda}\left(\frac{1}{1+\frac{1}{\lambda}}\right)^k.$$

The probability that there will be l or more events occurring at a point is given by

$$Pr\{e \geq l\} = 1 - \sum_{k=0}^{l-1} p_k.$$

We are interested in finding a threshold for distinguishing correct from random interpretations. This can be done by setting the threshold, f_0, to be the fraction of model features such that $l = mf_0$. If we choose a value δ for the probability that there will be mf_0 or more events occurring at random (i.e. a limit on the false positive rate), then the condition on f_0 is given by

$$1 - \sum_{k=0}^{mf_0-1} p_k \leq \delta.$$

Substituting for p_k yields

$$1 - \frac{1}{1+\lambda} \sum_{k=0}^{mf_0-1} \left(\frac{1}{1+\frac{1}{\lambda}}\right)^k \leq \delta$$

and using the geometric series relationship yields

$$1 - \frac{1}{1+\lambda} \left[\frac{\left(\frac{1}{1+\frac{1}{\lambda}}\right)^{mf_0} - 1}{\frac{1}{1+\frac{1}{\lambda}} - 1} \right] \leq \delta.$$

We can isolate f_0 by appropriate algebra:

$$f_0 \geq \frac{\log\left(\frac{1}{\delta}\right)}{m \log\left(1+\frac{1}{\lambda}\right)} \tag{3}$$

where

$$\lambda = ms\bar{c}.$$

The value of $\bar{c}$ depends on the particular type of feature being matched and the bounds on the sensor error. In the case of two dimensional edge fragments considered above, we derived

$$\bar{c} = \frac{2\epsilon_a \left[2\epsilon_p \bar{\alpha} \bar{L} + \pi \epsilon_p^2\right]}{2\pi D^2} = \frac{\epsilon_a}{\pi} \left[2\bar{\alpha} \frac{\epsilon_p}{D} \frac{\bar{L}}{D} + \pi \left(\frac{\epsilon_p}{D}\right)^2 \right].$$

Note that equation (3) exhibits expected behavior. If the noise in the data increases, then $\bar{c}$ increases, and so does the bound on f_0. Similarly, as the amount of occlusion increases, then so does $\bar{c}$ and thus the bound on f_0. As either m or s increases so does the bound on f_0, and as δ decreases f_0 increases.

Also note that for large values of ms, the logarithm in the denominator can be approximated by its first order term, and one gets the following approximation

$$f_0 \geq \left(\log \frac{1}{\delta}\right) s\bar{c}.$$

Thus, in the limit, the bound on the fraction of the model is linear in the number of sensory features, linear in the average size of the volumes in transformation space, and varies logarithmically with the inverse probability of a false match.

The expression for f_0 in equation (3) can yield values that are greater than 1.0, which makes no sense as a *fraction* of the model features. When f_0 is greater than 1.0 it means that for the given number and type of features, and the given bounds on sensor error, it is not possible to limit the probability of a false match to the chosen δ (even if all the model features are matched to some sensor feature).

Thus to obtain a value for the fraction of model features that must be matched in order to limit the probability of a random conspiracy to δ, we simply need to compute $\bar{c}$ for the particular parameters of our recognition task, and then use equation (3) to compute f_0.

There are several possible choices for δ. One could simply set δ to be some small number, e.g. $\delta = .001$ so that a false positive is likely to arise no more than

one time in a thousand. One could also set δ as a function of the scene complexity, e.g. some multiple of the inverse of the total number of data model pairings

$$\frac{\beta}{ms},$$

where β is an arbitrarily chosen constant.

A third possibility is to set δ so that the likelihood of a false positive, integrated over the entire transformation space, is small (e.g. less than 1). The idea is to determine the appropriate value of δ such that the expectation is that no random matches will occur. If we let ν be a measure of the sensitivity of our system in distinguishing transformations, then we could choose δ as

$$\delta = \frac{\nu}{2\pi D^2}.$$

For example, we could set ν to be a function of the noise in the data measurements, given by the uncertainty in orientation times the uncertainty in position: $(2\epsilon_a)(\pi\epsilon_p^2)$. In this particular case, we get

$$f_0 \geq \frac{\log\left(\frac{D^2}{\epsilon_a \epsilon_p^2}\right)}{m \log\left(1 + \frac{1}{ms\hat{c}}\right)}. \qquad (3a)$$

To illustrate the values for f_0, we graph representative examples in Figures 3–5. Figure 3 displays graphs of f_0 as a function of s, with $m = 32, c = .0002215$ (these numbers are taken from the recognition systems analyzed in section 5). Each graph is for a different value of δ. Note that as s gets large, the graphs become linear, as expected.

Figure 4 displays f_0 as a function of m for different values of δ. Here, $s = 100, c = .0002215$. Note that as expected, when m becomes large, f_0 becomes a constant independent of m.

Figure 5 displays f_0 as a function of the sensor error, for different values of δ. Here, $s = 100, m = 32$. The percentage of error along the horizontal axis p is used to define sensing errors of $\epsilon_a = p\pi$ and $\epsilon_p = p\overline{L}$. As expected, the threshold on f_0 increases with increasing error.

Allowing for weighted votes

The preceding analysis treated each data-model feature pairing equally, and bounded the probability that l such pairings would be consistent at random. Another approach is to weight the contribution of each data-model pairing by some measure. One common scheme is to use the size of each data feature as a weight. In the case of two dimensional edges, for example, a data-model pairing (j, J) would carry a weight of ℓ_j (the length of the data edge), so that transformations consistent with pairings of long data edges to model edges would be more highly valued than those involving short data edges.

We can modify our preceding analysis to handle this case as well. Note that the parameter λ essentially measures the average "vote" at each point in the transformation space. Since we have assumed that each volume of transformations consistent

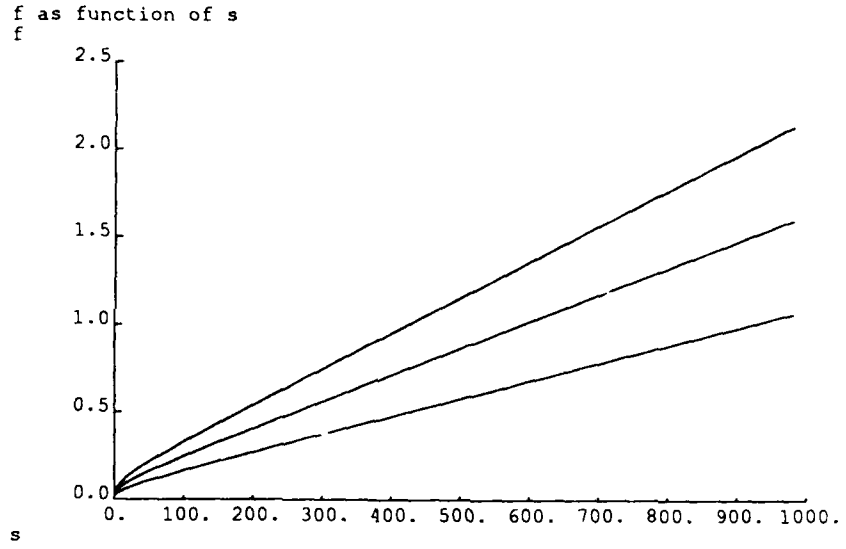

Figure 3. Graphs on bounds on threshold. f_0 is graphed as a function of s, with other parameters fixed. The three graphs are for $\delta = .0001, .001, .01$ from top to bottom respectively.

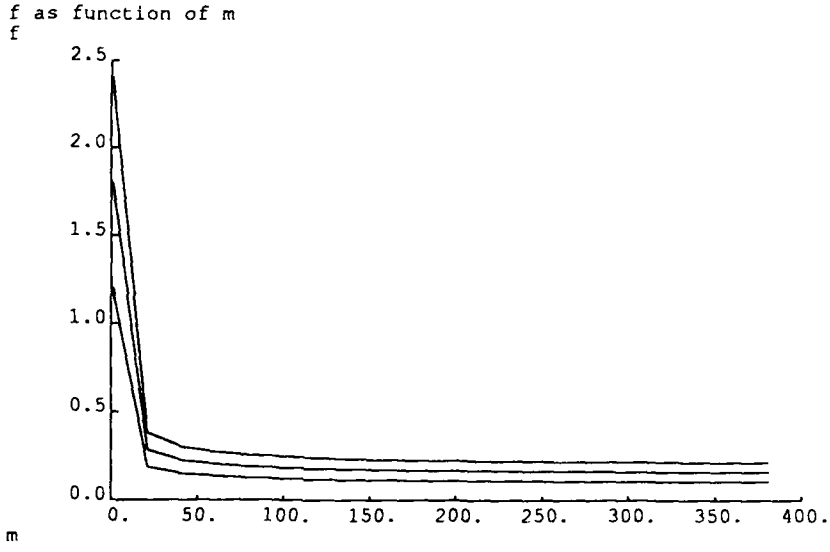

Figure 4. Graphs on bounds on threshold. f_0 is graphed as a function of m, with other parameters fixed. The three graphs are for $\delta = .0001, .001, .01$ from top to bottom respectively.

with some data-model feature pairing is independent, we can derive the expected weighted "vote" at any point in transformation space. As one might expect, due to

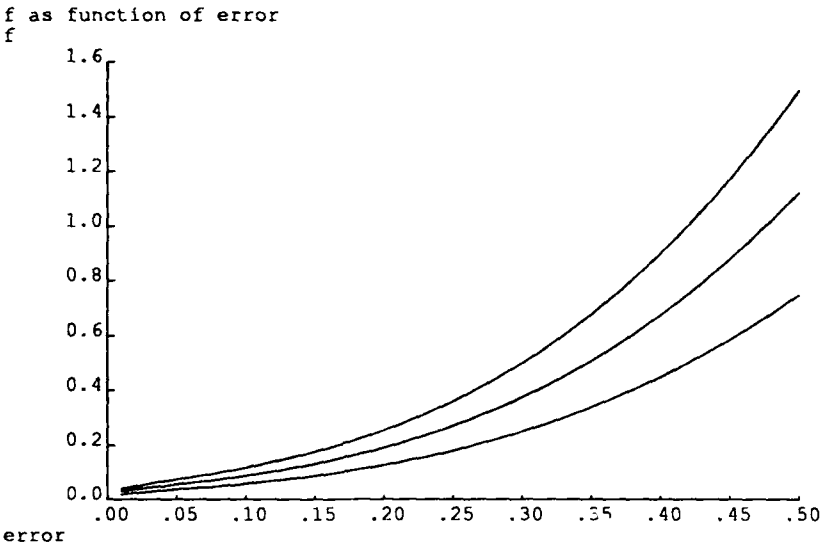

Figure 5. Graphs on bounds on threshold. f_0 is graphed as a function of error, with other parameters fixed. The three graphs are for $\delta = .0001, .001, .01$ from top to bottom respectively. The percentage of error along the horizontal axis p is used to define sensing errors of $\epsilon_a = p\pi$ and $\epsilon_p = p\overline{L}$.

the independence, this simply yields

$$\lambda = ms\bar{\ell}\bar{c}$$

where $\bar{\ell}$ is the average length of the data edges. Note that this is the average length over all data edges, not just those that match the object.

In this case we are interested in bounds on f_0 such that

$$1 - \sum_{k=0}^{m\overline{L}f_0 - 1} p_k \leq \delta.$$

Working through the same algebra as in the previous section leads to the following bound:

$$f_0 \geq \frac{\log\left(\frac{1}{\delta}\right)}{m\overline{L}\log\left(1 + \frac{1}{ms\bar{\ell}\bar{c}}\right)}. \tag{4}$$

5. Some Real World Examples

To demonstrate the utility of our method, in this section we analyze some working recognition systems that utilize a threshold on the fraction of model features which must be accounted for by a match. We find that the analysis predicts thresholds that are close to those that were determined experimentally. This suggests that the technique can be profitably used to analytically determine thresholds for model-based matching. Because our analysis shows that the proper threshold *varies* with

the number of model and data features, it is important to be able to set the threshold as a function of a particular matching problem rather than setting it once based on experimentation.

As a first example, we consider the application of the interpretation tree method [Grimson and Lozano-Pérez, 1984, 1987; Ettinger, 1987, 1988; Murray, 1987a, 1987b; Murray and Cook, 1988] to recognizing sets of two dimensional parts. In this approach, a tree of possible matching model and image features is constructed. Each level of the tree corresponds to one of the image features. At every node of the tree there is a branch corresponding to each of the model features, plus a special branch that accounts for model features that do not match the image. A path from the root to a leaf node maps each image feature onto some model feature or the special "no-match" symbol. The tree is searched by maintaining pairwise consistency among the nodes along a path. Consistency is checked using distance and angle relations between the model and image features specified the nodes. If a given node is inconsistent with any node along the path to the root then the subtree below that point is pruned from further consideration.

A consistent path from the root to a leaf that accounts for more than some fraction of the model features is accepted as a correct match. This threshold is chosen experimentally. In our analysis of thresholds for the interpretation tree method, we use the parameters for the objects demonstrated in [Grimson and Lozano-Pérez 1987], and the parameters for a typical scene in the experimentation described there. These values are substituted into equation (2), and then a threshold f_0 is computed using equations (3) and (3a).

In the experiments reported in [Grimson and Lozano-Pérez, 1987], the following parameters hold:

$$\begin{aligned} m &= 32 \\ s &= 100 \\ \overline{L} &= 23.959 \\ \epsilon_p &= 10 \\ \epsilon_a &= \frac{\pi}{10}. \end{aligned}$$

We have computed $\overline{c}$ as a function of the amount of occlusion $\overline{\alpha}$, and then determined the corresponding threshold f_0 on the fraction of model features. Note that an occlusion of 1 represents the limiting case in which only a point on the line is visible. The results are given in Table 1. The first column of the table shows the values of f_0 computed using equation (3a). Recall that this method integrates over the transformation space in order to limit the expectation of a randomly occurring match by setting

$$\frac{1}{\delta} = \left(\frac{D^2}{\epsilon_a \epsilon_p^2}\right).$$

For comparison, the second and third columns of the table are computed using equation (3), with the probability of a random match, δ, set to .001 and .0001,

Occlusion	f, eqn (3a)	f, with $\delta = .001$	f, with $\delta = .0001$
0.0	0.225	0.173	0.230
0.1	0.244	0.188	0.250
0.2	0.263	0.202	0.270
0.3	0.282	0.217	0.289
0.4	0.301	0.231	0.308
0.5	0.319	0.245	0.327
0.6	0.337	0.259	0.346
0.7	0.355	0.273	0.364
0.8	0.374	0.287	0.383
0.9	0.392	0.301	0.401
1.0	0.409	0.315	0.420

Table 1. Predicted bounds on termination threshold, as a function of the amount of occlusion, for trials of the **RAF** system.

respectively.

As expected, the bound on f increases as the amount of occlusion increases. Note that this bound is limited in scope even as the occlusion factor ranges over the entire possible range, that is, even for occlusions ranging from none 0 to all 1, the bound on f only varies over a range of 0.225 to 0.409. It is interesting to compare these results with empirical observations. Grimson and Lozano-Pérez report that in running the **RAF** system on a variety of images of this type using thresholds of $f = .4$ resulted in no observed false positives, while using thresholds of $f = .25$ would often result in a few false positives. Since on average the occlusion was roughly .5, this observation fits nicely with the predictions of Table 1, namely that a threshold of .4 should yield no errors, while a threshold of .25 cannot guarantee such success.

If we use the lengths of the data features to weight the individual feature matchings then substituting into equation (4) leads to the predictions shown in Table 2. These values were computed using equation (3a) in the same manner as the first column of Table 1. Again, this agrees with empirical experience for the **RAF** system, in which weighted matching using thresholds of $f = .25$ almost always led to no false positives, while using thresholds of $f = .10$ would often result in a few false positives.

As a second example, we consider the **HYPER** system of Ayache and Faugeras [1986]. Similar to **RAF**, **HYPER** also uses geometric constraints to find matches of data to models. An initial match between a long data edge and a corresponding model edge is used to estimate the transformation from model coordinates to data coordinates. This estimate is then used to predict a range of possible positions for unmatched model features, and the image is searched over this range for potential matches. Each potential match is evaluated using position and orientation constraints, and the best match within error bounds is added to the current interpretation. The additional model-data match is used to refine the estimate of the

Occlusion	f with $\overline{\ell} = \overline{L}$	f with $\overline{\ell} = .75\overline{L}$	f with $\overline{\ell} = .5\overline{L}$
0.0	0.119	0.091	0.062
0.1	0.136	0.103	0.071
0.2	0.153	0.116	0.079
0.3	0.171	0.129	0.088
0.4	0.188	0.142	0.097
0.5	0.205	0.155	0.105
0.6	0.222	0.168	0.114
0.7	0.240	0.181	0.123
0.8	0.257	0.194	0.131
0.9	0.274	0.207	0.140

Table 2. Predicted bounds on termination threshold, as a function of the amount of occlusion, for trials of the **RAF** system. In this case, the lengths of the matched edges is used, instead of just the number of matched edges.

transformation, and the process is iterated.

Although not all of the parameters needed to apply our analysis are given in the paper, we can estimate many of them from the illustrations provided in the article. Given several estimates for the error in the measurements, a range of values for the threshold f are listed in Table 3. Object-1 and Object-2 refer to the object labels used by Ayache and Faugeras. In these examples, we use orientational errors of $\epsilon_a = \pi/10$ and $\pi/15$ radians and positional errors of $\epsilon_p = 3$ pixels.

In **HYPER**, a threshold of .25 is used to discard false positives, and Ayache and Faugeras report the observation of no false positives during a series of experiments with their system. For the two objects listed in Table 3, Ayache and Faugeras report that their system found interpretations of the data accounting for a fraction of .55 of the model for Object-1 and accounting for a fraction of .40 of the model for Object-2. Both these observations are in agreement with the thresholds predicted in Table 3, for different estimates of the data error.

Thus for two different recognition systems (**RAF** and **HYPER**), using both weighted and unweighted matching schemes, we see that the technique developed in this paper yields matching thresholds that are similar to those determined experimentally by the designers of the systems.

6. Conclusion

In order to determine what constitutes an acceptable match of a model to an image, most recognition systems use an empirically determined threshold on the fraction of model features that must be accounted for. In this paper we have developed a technique for analytically determining the fraction of model features f_0 that must be matched in order to limit the probability of a random conspiracy of the data to some level δ. This fraction f_0 is a function of the type of feature, the number of

Occlusion	Object-1		Object-2	
	f, ($\epsilon_a = \frac{\pi}{10}$)	f, ($\epsilon_a = \frac{\pi}{15}$)	f, ($\epsilon_a = \frac{\pi}{10}$)	f, ($\epsilon_a = \frac{\pi}{15}$)
0.0	0.224	0.185	0.206	0.168
0.1	0.243	0.199	0.225	0.181
0.2	0.261	0.212	0.243	0.195
0.3	0.279	0.225	0.262	0.208
0.4	0.297	0.238	0.280	0.221
0.5	0.315	0.251	0.298	0.234
0.6	0.333	0.264	0.316	0.247
0.7	0.350	0.277	0.335	0.260
0.8	0.368	0.289	0.353	0.273
0.9	0.386	0.302	0.371	0.285

Table 3. Predicted bounds on termination threshold, as a function of the amount of occlusion, for trials of the HYPER system. The first two columns for f are Object-1, the final two for Object-2.

model features, m, the number of sensor features, s and bounds on the translation error ϵ_p and the angular error ϵ_a of the sensor and feature detector.

Our analysis shows that the proper threshold varies with the number of model and data features. A threshold that is appropriate for relatively few data features is not appropriate when there are many data features. Thus it is important to be able to set the threshold as a function of a particular matching problem, rather than setting a single threshold based on some experimentation. The technique developed in this paper provides a straightforward means of computing a matching threshold for the values of m and s found in a given recognition situation.

We have applied the technique to two existing recognition systems, and found that the predicted thresholds are close to those that were determined experimentally. This suggests that the method can be profitably used to analytically determine thresholds for model-based matching systems.

Appendix: Extending the method to other cases

So far, we have demonstrated our method on the case of recognition of rigid two-dimensional objects from two-dimensional edges or vertices. We can readily extend our method to other cases as well. In general, equations (3) and (4) still hold, with the proviso that $\bar{c}$ changes as the problem changes. First we consider adding scaling to the two-dimensional transformation, and then we consider three-dimensional transformations.

Objects that scale

First, we consider the case in which a two-dimensional object is free to scale within

some predefined range. In this case, the space of possible transformations is four-dimensional, having two dimensions for translation parameters, one for rotation, and one for scale.

In this case, the transformation from model coordinates to sensor coordinates may be represented by

$$\mathbf{v}_s = sR_\theta \mathbf{V}_M + \mathbf{V}_0$$

where $\mathbf{V}_M$ is a vector in model coordinates, R_θ is a rotation matrix corresponding to an angle of θ, s is a scale factor, $\mathbf{V}_0$ is a translation offset, and $\mathbf{v}_s$ is the corresponding vector in sensor coordinates.

As in the earlier case, if we consider the conditions on the transformation so that the endpoint of a data edge corresponds to the endpoint of a model edge, we find that

$$\mathbf{m}_j - \frac{\ell_j}{2}\hat{\mathbf{t}}_j = sR_{\theta_m}\left[\mathbf{M}_J - \frac{L_J}{2}\hat{\mathbf{T}}_J\right] + \mathbf{V}_0$$

We also have the condition that

$$sL_J \geq \ell_j$$

so that

$$s \geq \frac{\ell_j}{L_J}$$

or if we allow for error in the measurements, that

$$s \geq \frac{\ell_j - 2\epsilon_p}{L_J}.$$

Hence, for each choice of s, the translation is

$$\mathbf{V}_0 = \mathbf{m}_j - sR_{\theta_m}\mathbf{M}_J + \frac{sL_J - \ell_j}{2}R_{\theta_m}\hat{\mathbf{T}}_J.$$

Similarly, if the other endpoints align, we get

$$\mathbf{V}_0 = \mathbf{m}_j - sR_{\theta_m}\mathbf{M}_J - \frac{sL_J - \ell_j}{2}R_{\theta_m}\hat{\mathbf{T}}_J.$$

Because any intermediate position is also acceptable, the set of translations consistent with matching model edge J to data edge j is given by

$$\mathbf{V}_0(\theta, s) = \left\{ \mathbf{m}_j - sR_{\theta_m}\mathbf{M}_J + \gamma R_{\theta_m}\hat{\mathbf{T}}_J \;\middle|\; \gamma \in \left[-\frac{sL_J - \ell_j}{2}, \frac{sL_J - \ell_j}{2}\right] \right\}. \quad (5)$$

Hence, matching model edge J to data edge j yields a set of points in transform space $\mathcal{P}$, with a single value for the rotation parameter and a set of values for the translation, that correspond to a line of length $sL_J - \ell_j$, with orientation $R_{\theta_m}\hat{\mathbf{T}}_J$ in the x–y plane, where s can range from

$$s \geq \frac{\ell_j - 2\epsilon_p}{L_J}$$

to some predefined maximum s_h.

To determine the full volume of transformation space consistent with a data-model feature pairing, we must allow for noise in the measurements. As in the non-scaled case, we can integrate over a range of orientations within ϵ_a of the computed

one, and we must also integrate along the line of translations defined above as s varies. This implies that the full volume is given by

$$2\epsilon_a \int_{s=\frac{\ell_j - 2\epsilon_p}{L_j}}^{s_h} \left(2\epsilon_p \left(sL_j - \ell_j\right) + \pi\epsilon_p^2\right) ds$$

which reduces to

$$2\epsilon_a \left(s_h - \frac{\ell_j - 2\epsilon_p}{L_j}\right) \left[\pi\epsilon_p^2 + 2\epsilon_p \left(\frac{s_h L_J - \ell_j}{2} - \epsilon_p\right)\right].$$

This is normalized by the total volume of transformation space

$$2\pi D^2 \left(s_h - 1\right)$$

to yield

$$c_s = \frac{\epsilon_a}{\pi} \left(\frac{s_h - \frac{\ell_j - 2\epsilon_p}{L_j}}{s_h - 1}\right) \left[\pi \left(\frac{\epsilon_p}{D}\right)^2 + 2\frac{\epsilon_p}{D} \left(\frac{s_h L_J - \ell_j}{2D} - \frac{\epsilon_p}{d}\right)\right].$$

For cases involving scale, we can substitute c_s in place of c in the earlier analysis.

Three dimensional case

As a second extension, consider the problem of recognizing three-dimensional objects from three-dimensional edges. In this case, the transformation space is six dimensional, with three dimensions for translational components, and three for rotational components. As in the previous cases, we must deduce an expression for c that holds in this case.

We begin with the rotational parameters. Given the unit tangent vector of a model edge and of a data edge, there is a set of rotations that will consistently map the model tangent into the data tangent. The axis of rotation that will accomplish this lies anywhere in the great circle on the unit sphere equidistant from the two tangent vectors. For each such axis of rotation, there is a unique angle of rotation that will effect the mapping. When we allow for error, the data tangent is only known to within a cone of radius ϵ_a, and hence the great circle expands into a band of feasible axes of rotation. If we integrate out the volume of feasible rotations, we get

$$2\int_{\theta=0}^{2\pi} \int_{\phi=\frac{\pi}{2}-\cos^{-1}\epsilon_a}^{\frac{\pi}{2}} \sin\phi \, d\phi \, d\theta = 4\pi\sqrt{1-\epsilon_a^2}.$$

To account for the translation, we find that an analysis similar to the two dimensional case holds. In particular, the set of feasible translations is a cylinder of radius ϵ_p of length αL, where α is the amount of occlusion of the edge, capped by two hemi-spheres of radius ϵ_p. Hence, the overall volume is given by

$$4\pi\sqrt{1-\epsilon_a^2} \left(\alpha L \pi \epsilon_p^2 + \frac{4}{3}\pi\epsilon_p^3\right).$$

If the linear range of values for each dimension of translation is D, then the normalized coefficient in the three dimensional case is given by

$$c = \frac{\sqrt{1-\epsilon_a^2}}{2}\left(\alpha\frac{L}{D} + \frac{4}{3}\left(\frac{\epsilon_p}{D}\right)\right)\left(\frac{\epsilon_p}{D}\right)^2.$$

References

Ayache, N. & O.D. Faugeras, 1986, "HYPER: A new approach for the recognition and positioning of two-dimensional objects," *IEEE Trans. Pat. Anal. Mach. Intel.*, **8** no. 1, pp. 44–54.

Besl, P.J. & R.C. Jain, 1985, "Three-dimensional object recognition," *ACM Computing Surveys*, **17** no. 1, pp. 75–154.

Chin, R.T. & C.R. Dyer, 1986, "Model-based recognition in robot vision," *ACM Computing Surveys*, **18** no. 1, pp. 67–108.

Clemens, D.T., 1986, The recognition of two-dimensional modeled objects-in images, M. Sc. Thesis, Massachusetts Institute of Technology, Electrical Engineering and Computer Science.

Ettinger, G.J., 1987, "Hierarchical Object Recognition Using Libraries of Parameterized Model Sub-parts," MIT AI Lab Technical Report 963.

Ettinger, G.J., 1988, "Large Hierarchical Object Recognitino Using Libraries of Parameterized Model Sub-parts," *IEEE Conf. on Comp. Vis. & Pattern Recog.*, pp. 32–41.

Faugeras, O.D. & M. Hebert, 1986, "The representation, recognition and locating of 3-D objects," *Int. J. Robotics Research*, **5** no. 3, pp. 27–52.

Feller, W., 1968, *An Introduction to Probability Theory and Its Applications*, New York, Wiley.

Fischler, M.A. & R.C. Bolles, 1981, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, **24**, pp. 381–395.

Grimson, W.E.L., 1989a, "On the recognition of curved objects in two dimensions," *IEEE Trans. Pat. Anal. Mach. Intel.*, **11** no. 6, pp. 632–643.

Grimson, W.E.L., 1989b, "On the Recognition of Parameterized 2D Objects," *International Journal of Computer Vision*, **2** no. 4, pp. 353–372.

Grimson, W.E.L., 1989c, "The Combinatorics of Object Recognition in Cluttered Environments Using Constrained Search," *Artificial Intelligence*, to appear.

Grimson, W.E.L. & D.P. Huttenlocher, 1989a, "On the Sensitivity of the Hough Transform for Object Recognition," *IEEE Pattern Analysis and Machine Intelligence*, to appear.

Grimson, W.E.L. & T. Lozano-Pérez, 1984, "Model-based recognition and localization from sparse range or tactile data," *Int. Journ. Robotics Res.*, **3** no. 3, pp. 3–35.

Grimson, W.E.L. & T. Lozano-Pérez, 1987, "Localizing overlapping parts by searching the interpretation tree," *IEEE Trans. Pat. Anal. Mach. Intel.*, **9** no. 4, pp. 469–482.

Huttenlocher, D.P. & S. Ullman, 1987, "Object recognition using alignment," *Proc. First Intern. Conf. Comp. Vision*, London, pp. 102–111.

Huttenlocher, D.P. & S. Ullman, 1988, "Recognizing Solid Objects by Alignment," *Proc. DARPA Image Understanding Workshop*, pp. 1114–1124.

Ikeuchi, K., 1987, "Generating an interpretation tree from a CAD model for 3d-object recognition in bin-picking tasks," *International Journal of Computer Vision*, **1** no. 2, pp. 145–165.

Murray, D.W., 1987a, "Model-based recognition using 3D structure from motion," *Image and Vision Computing*, pp. 85–90.

Murray, D.W., 1987b, "Model-based recognition using 3D shape alone," *Computer Vision, Graphics and Image Processing*, **40**, pp. 250–266.

Murray, D.W. & D.B. Cook, 1988, "Using the orientation of fragmentary 3D edge segments for polyhedral object recognition," *Intern. Journ. Computer Vision*, **2** no. 2, pp. 153–169.

Stockman, G., 1987, "Object recognition and localization via pose clustering," *Comp. Vision, Graphics, Image Proc.*, **40**, pp. 361–387.

Stockman, G., S. Kopstein & S. Benett, 1982, "Matching images to models for registration and object detection via clustering," *IEEE Trans. Pattern Anal. and Machine Intel.*, **3** no. 3, pp. 229–241.

Thompson, D.W. & J.L. Mundy, 1987, "Three-dimensional model matching from an unconstrained viewpoint," *Proceedings of the International Conference on Robotics and Automation*, Raleigh, NC, pp. 208–220.